5-Hydroxymethylcytosine Remodeling Precedes Lineage Specification during Differentiation of Human CD4+ T Cells

5-hydroxymethylcytosine remodeling precedes lineage specification during human T-cell differentiation and marks disease-associated regulatory regions

Authors/Affiliations

Colm E. Nestor,[1,7*] Antonio Lentini,[1,7] Cathrine Hägg Nilsson,[1] Danuta Gawel,[1] Mika Gustafsson,[1] Lina Mattson,[1] Hui Wang,[2] Olof Rundquist,[1] Richard R. Meehan,[3] Bernward Klocke,[4] Martin Seifert,[4] Stefanie M. Hauck,[5] Helmut Laumen,[6] Huan Zhang,[1,7] Mikael Benson[1,7*]

1. Centre for Individualized Medicine, Department of Pediatrics, Faculty of Medicine, Linköping University, Linköping, Sweden.

2. MD Anderson Cancer Centre, Houston, Texas, USA

3. MRC Human Genetics Unit, MRC Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh EH4 2XU, Scotland

4. Genomatix Software GmBH, 80335 Munich, Germany

5. Research Unit Protein Science, Helmholtz Zentrum München, German Research Center for Environmental Health GmbH, 85764 Neuherberg, Germany

6. Else Kröner-Fresenius-Center for Nutritional Medicine, Chair of Nutritional Medicine, ZIEL, DZD, Clinical Cooperation Group at the Helmholtz Zentrum München, Klinikum rechts der Isar, 81675 Munich, Germany, and Technische Universität München, 85350 Freising-Weihenstephan,, Germany

7. These authors contributed equally to this work

Contact

*Corresponding authors: colm.nestor@liu.se (C.E.N.), mikael.benson@liu.se (M.B.)

Summary

Conversion of 5-methylcytosine (5mC) to 5-hydroxymethylcytosine (5hmC) by TET enzymes is critical to DNA methylation reprogramming. Mutation of *TET* genes in T-cell malignancies and loss of TET activity resulting in auto-immune phenotypes in mouse, suggests important roles in T-cell biology. Consistent with this, we report early and widespread 5mC/5hmC remodeling during human $CD4^+$ T-cell differentiation *ex vivo* at genes and cell-specific enhancers with known T-cell function. We observed similar DNA de-methylation in $CD4^+$ memory T-cells *in vivo,* indicating that early remodeling events persist in differentiated cells. Underscoring their important function, 5hmC loci were highly enriched for genetic variants associated with T-cell diseases and T-cell-specific chromosomal interactions. Extensive functional validation of 22 risk variants revealed novel, potentially pathogenic, mechanisms in diabetes and multiple sclerosis. Our results support 5hmC-mediated DNA de-methylation as a key component of $CD4^+$ T-cell biology in humans, with important implications for identifying disease-associated genetic variants.

Highlights

- 5hmC remodeling is widespread during human $CD4^+$ T-cell differentiation
- Early 5hmC gains predict loss of DNA methylation in differentiated cells
- 5hmC remodeling *in vitro* predicts loss of DNA methylation *in vivo*
- 5hmC loci are enriched for functional, T-cell disease associated genetic variants

eTOC

Nestor, Lentini and colleagues reveal widespread 5hmC-mediated DNA de-methylation during *in vitro* differentiation of human $CD4^+$ T-cells and propose 5hmC profiling as a novel and effective approach for the identification of regulatory variants in human disease.

INTRODUCTION

Differentiation of $CD4^+$ T-cells into effector or regulatory subtypes is critical to adaptive immunity. Upon contact with antigens, T-cells differentiate into various T-helper (Th) cell subsets, such as Th1, Th2, Th17 or regulatory T (Treg) cells (Yamane and Paul, 2013), which mediate or inhibit immune responses. Inappropriate $CD4^+$ T-cell differentiation is associated with several autoimmune and inflammatory diseases, including rheumatoid arthritis (RA), psoriasis, allergy, asthma, multiple sclerosis (MS) and type 1 diabetes (Gustafsson et al., 2015; Licona-Limon et al., 2013; Wahren-Herlenius and Dorner, 2013). The lack of a strong genetic component and increasing prevalence of these diseases suggests an epigenetic contribution to their pathogenesis, and changes in T-cell DNA methylation patterns have been reported in MS, allergy, and RA (Graves et al., 2013; Liu et al., 2013; Nestor et al., 2014a).

Appropriate differentiation of Th subsets requires widespread remodeling of the T-cell epigenome, including DNA de-methylation of many master regulators of the differentiation process, such as *Il4, Il5, Il13* (Th2), *Ifng* (Th1) and *FOXP3* (Treg) (Janson et al., 2011; Lee et al., 2006). 5-hydroxymethylcytosine (5hmC) was recently discovered to be highly abundant in the human genome and generated by hydroxylation of 5-methylcytosine (5mC) by members of the Ten-Eleven-Translocation (TET1/2/3) family of enzymes (Tahiliani et al., 2009). 5hmC can subsequently be resolved to unmodified cytosine, completing the process of DNA de-methylation (**Figure S1A**). Significantly, *TET* loss-of-function mutations have been identified in several hematological malignancies, with the highest frequency in adult $CD4^+$ T-cell cancers (Kalender Atak et al., 2012; Lemonnier et al., 2012). Moreover, *Tet2* knockout mice exhibit impaired differentiation of hematopoietic stem cells and developed autoimmune phenotypes (Ichiyama et al., 2015; Ko et al., 2011; Li et al., 2011; Yang et al., 2015). Despite the valuable insights into the role of TET-5hmC during differentiation of mammalian $CD4^+$ T-cells obtained from mouse models (Ichiyama et al., 2015; Ko et al., 2011; Tsagaratou et al., 2014; Yang et al., 2015), little is known about the importance of DNA de-methylation in human $CD4^+$ T-cell differentiation and its contribution to the pathogenesis of complex immune diseases.

We generated genome-wide maps of 5hmC, 5mC and gene expression during early and late stages of human $CD4^+$ T-cell differentiation *ex vivo*. Changes in 5hmC were widespread during both activation and differentiation of $CD4^+$ T-cells, occurred at a variety of activating and repressive regulatory elements, and coincided with tight regulation of *TET* gene expression. Significantly, all early 5hmC and 5mC remodeling occurred in the complete absence of replication, suggesting an 'active', enzymatic remodeling mechanism. Using genetic overexpression we showed that tight regulation of *TET* levels was required for appropriate expression of key lineage specific transcription factors and cytokines. We confirmed these

findings *in vivo* by transcriptional and epigenetic profiling of human naïve $CD4^+$ T cells (NT), central memory (T_{CM}) and effector memory T-cells (T_{EM}). Supporting the disease relevance of 5hmC-mediated DNA de-methylation, loci gaining 5hmC during early T-cell differentiation were highly enriched for variants associated with T-cell related diseases at a diversity of *cis*- and *trans*- gene regulatory elements. Moreover, these regions were also enriched for T-cell specific chromosomal interactions, supporting their importance in T-cell biology. We undertook further functional characterization of the effects of over 20 predicted regulatory variants on the level of DNA-protein interactions, and reveal novel, potentially pathogenic, mechanisms in diabetes and multiple sclerosis. Our results support 5hmC-mediated DNA de-methylation as a key component of $CD4^+$ T-cell biology in humans, and 5hmC profiling as a novel and cost-effective approach for identification of regulatory genetic variants in complex immune disease.

RESULTS

5hmC remodeling during $CD4^+$ T-cell differentiation occurs in absence of replication and is enriched at key regulatory genes

To dissect the role of DNA de-methylation in human $CD4^+$ T cell function, we took advantage of the ability to differentiate pure human naïve T-cells into T helper cell subsets *in vitro* (**Figure 1A**). This powerful approach allowed direct observation of the early 5hmC remodeling events, occurring in the absence of DNA replication, that result in the stable lineage-specific 5mC profiles observed in differentiated T helper cell subsets. Appropriate differentiation into Th1 and Th2 lineages was confirmed by gene expression microarray and qRT-PCR of key lineage-specific genes (**Figures S1B-C & Table S1**). *In vitro* polarization of NT cells resulted in a >5-fold loss ($P_{T\text{-}TEST} < 0.05$) of global 5hmC levels (**Figure 1B**). A small, but consistent 10-15% loss of global 5hmC levels was also observed after only 1 day of *in vitro* polarization, consistent with changes in the 5-hydroxymethylome occurring before the onset of DNA replication

(**Figures 1B & S1D-E**). Next, we combined 5hmC DNA immuno-precipitation with massively parallel sequencing (hMeDIP-seq) to generate the first hydroxymethylomes of human $CD4^+$ T-cells differentiating into Th1 and Th2 effector cells in two replicate experiments. 5hmC profiles were generated in naïve cells and both at early (1 day) and late (5 day) time-points during polarization, allowing identification of both early replication-independent changes as well as later subset-specific changes in 5hmC. We generated 1 billion paired reads in total and for each condition 35-60 million paired reads were uniquely mapped to the human genome (hg19) (**Table S2**). 5hmC showed characteristic enrichment in gene bodies and depletion at transcription start sites as previously reported in other cell types (**Figure 1C**) and a clear association with actively transcribed genes (**Figure 1D**) (Nestor et al., 2012; Song et al., 2011). Similar to results reported in mouse $CD4^+$ T-cells, the distribution of 5hmC enrichment was significantly overrepresented in genic regions in all T-cell subsets (**Figure 1E & Table S2**) (Ichiyama et al., 2015; Tsagaratou et al., 2014).

To understand the biological relevance of these changes, we mapped the 1,000 largest gains and losses of 5hmC to their nearest gene and subjected these to gene ontology (GO) enrichment analysis. Strikingly, whereas regions losing 5hmC were not enriched in any biological processes, regions gaining 5hmC at day 1 in either Th1 or Th2 cells were highly significantly ($P_{ADJUSTED} < 1 \times 10^{-8}$) enriched for genes associated with T-cell activation (**Figure 1F and data not shown**), including the signature genes *IFNG*, *TBX21*, *FOXP3*, *ZAP70*, *IL2RA*, *IL7R*, *IRF4*, *CD5*, *AIM2*, *CCR2*, *CCR5*, *IL1R2*, *IL26* and *IL32R* (**Figure 1G and data not shown**). Globally, large-scale remodeling of 5hmC had occurred by day 1, with over 10,000 and 5,000 regions significantly gaining or losing 5hmC, respectively (**Figure S1F and Table S2**). The scale of change was far greater in regions gaining 5hmC (**Figure S1G**).

Taken together, these results reveal genome-wide, replication-independent re-programming of 5hmC during $CD4^+$ T-cell differentiation at genes key to appropriate T-cell activation and differentiation.

Lineage specification of human $CD4^+$ T-cells is preceded by 5hmC-mediated DNA de-methylation of gene regulatory elements

To relate 5hmC remodeling during differentiation to changes in DNA methylation (5mC), we subjected the same *in vitro* differentiated human Th1 and Th2 cells to DNA methylation profiling using Infinium 450K methylation arrays. Although not genome-wide, these arrays provide quantitative, base-resolution methylation measurements at ~450,000 CpG sites throughout the genome, allowing accurate detection of even small (> 3%) changes in absolute methylation levels. We observed a clear bias towards loss of 5mC in both Th1 and Th2 differentiated cells with more pronounced changes occurring by day 5 (**Figure 2A**). However, some changes in DNA methylation were also observed at day one (123 CpGs >10% loss, $P < 0.05$), in complete absence of replication (**Figures S2A & Table S3**) suggesting that such changes are active, and not secondary to DNA replication. Importantly, regions gaining 5hmC at day one of differentiation in both Th1 and Th2 cells were highly significantly enriched for CpGs losing 5mC at day 5 ($P < 0.0001$, Fisher exact test) (**Figures 2B & S2B**), suggesting that DNA de-methylation at these loci occurs primarily via a 5hmC-mediated process. Indeed, loci showing the greatest loss of 5mC during Th1 and Th2 differentiation were also those showing the greatest gains in 5hmC after only 1 day of polarization (**Figure 2C**). Finally, genes showing loss of DNA methylation (> 30% loss of 5mC) in their promoters during differentiation were enriched for functional terms related to the immune response (**Figure S2C**), further indicating that DNA de-methylation occurs in a targeted manner. Taken together, these findings suggest a model whereby loci to be de-methylated in differentiated T-cell subtypes first undergo

enzymatic hydroxylation of 5mC to 5hmC followed by both replication-independent excision and replication-dependent dilution of 5hmC during successive rounds of DNA replication.

As the majority (> 95%) of loci undergoing 5hmC remodeling during early $CD4^+$ T-cell differentiation occurred outside annotated gene promoters, we hypothesized that these loci might represent gene regulatory elements. Indeed the TET enzymes interact with numerous chromatin modifiers (HDAC1, HDAC2, EZH2, SIN3A) and co-repressor complexes (NuRD) which may serve to target TET methylcytosine dioxygenase activity to regulatory elements within the genome (Delatte et al., 2014) and elevated 5hmC levels have been observed at active enhancers in several systems (Lu et al., 2014; Tsagaratou et al., 2014). Using published chromatin-immunoprecipitation (ChIP)-sequencing data, we analyzed the association between 5mC/5hmC remodeling in $CD4^+$ T-cells and the enhancer-associated histone modifications, H3K4me1 and H3K27ac (Hawkins et al., 2013) (**Figure 2D**). Indeed, Th1- and Th2-specific active enhancer elements marked by H3K27ac (Hawkins et al., 2013) were highly enriched for regions showing 5mC loss during differentiation (~15-fold, $p < 0.001$) (**Figures 2E-G**).

Ectopic expression of *TET1* results in dysregulation of key cytokine and chemokine genes during differentiation

Expression profiling by qPCR showed that *TET1,* in contrast to *TET2* and *TET3,* was highly expressed in T-cells, as well as thymus compared to other human tissues and immune cells (**Figure 3A & Figure S3A**). Moreover, whereas the absolute levels of *TET1* were lower than those of *TET2* and *TET3* during T-cell differentiation, *TET1* alone underwent, rapid and stable down-regulation during early T-cell activation (>10-fold reduction; $P < 0.01$, t-test) (**Figures 3B-3C & Figure S3B**). Indeed, *TET1* silencing was observed as early as 6 hours after initiation of differentiation (**Figure S3B**). Unlike Tet2 and Tet3, Tet1 has been shown to bind to Polycomb target gene promoters and associate with chromatin repressors, thereby having a role

in direct transcriptional repression unrelated to its enzymatic activity (Williams et al., 2011; Wu et al., 2011). Thus, we sought to dissect the functional role of TET1 during early T-cell differentiation by ectopic expression of full-length human TET1 (TET1fl) (Tahiliani et al., 2009), the catalytic domain of TET1 (TET1cd) or a mutated catalytic domain of TET1 (TET1mut) lacking enzymatic activity (Guo et al., 2011) (**Figures 3D and S3C-D**). Notably, overexpression of TET1fl led to significant dysregulation of chemokine and cytokine gene mRNA levels ($log_2FC > 0.5$), including several key regulators of Th1/Th2 differentiation, including *IFNG*, *IL12RB2*, *HAVCR2*, *GATA3* and *IL5*. (**Figures 3E and S3E**). Far fewer changes in gene expression were observed with TET1cd (**Figures 3F and S3F, Table S4**), the majority of which were associated with calcium ion channel activity (**Figure S3F**). These findings suggests that TET1 might act as a direct transcriptional repressor during T-cell differentiation. However, as the transcriptional programs of each T helper subset generally inhibit those of the other lineages future studies are warranted to fully elucidate the functional role of TET1. Nevertheless, the results clearly support an important role for TET1 in the regulation of key genes during lineage-specification of T-helper cell subsets.

5hmC remodeling during early differentiation of human CD4$^+$ T-cells predicts loss of 5mC in CD4$^+$ memory T-cells *in vivo.*

We and other have previously shown that adaptation of primary mammalian cells to culture can affect the genomic distribution of 5hmC (Nestor et al., 2012; Nestor et al., 2015). Thus, having determined the DNA methylation dynamics of human T-cell differentiation *in vitro*, we sought to establish if similar changes occurred *in vivo*. After activation and differentiation of naïve T-cells (NT), a small proportion of cells remain as long-lived memory cell populations, which can be sub-divided into central memory (T_{CM}) and effector memory (T_{EM}) subsets based on their function and homing capacity (Sallusto et al., 2004). We isolated primary human NT

($CD4^+CD45RO^-CCR7^+$), T_{CM} ($CD4^+CD45RO^+CCR7^+$) and T_{EM} ($CD4^+CD45RO^+CCR7^-$) cells by FACS (purity >95%) for analysis (**Figure S4A**). Consistent with our findings in polarized T-cells, *in vivo* memory subsets exhibited decreased ($P < 0.05$, t-test) levels of *TET1* and *TET3* (**Figure 4A**) as well as lower global 5hmC content (**Figure 4B**), verifying that TET inactivation events observed during polarization also occur *in vivo*. Next, we generated gene expression and genome-wide DNA methylation profiles for NT, T_{CM} and T_{EM} cells. Gene expression of T_{CM} and T_{EM} cells showed distinct profiles, with T_{EM} cells expressing high levels of effector molecules such as *IL4*, *IFNG* and *CSF2*, while T_{CM} cells showed an intermediate profile, both in expression levels and number of genes up- or downregulated (**Figure 4C, S4B-C & Table S5**), consistent with a linear differentiation program of NT → T_{CM} → T_{EM}. Interestingly, DNA methylation profiles could clearly separate NT, T_{CM} and T_{EM} cells indicating that each subset is epigenetically distinct (**Figure 4D & S4D)**. Furthermore, T_{EM} cells showed extensive genome-wide de-methylation, exceeding that of T_{CM} (**Figure 4E**), consistent with the notion that T_{EM} cells are more terminally differentiated. As expected, DNA methylation changes in gene promoters was correlated with gene expression (*rho* = -0.43, p = 5.9e-08) (**Figure 4F**). Having characterized methylome- and transcriptome dynamics in memory T-cells, we sought to relate these changes to early hydroxymethylome reprogramming during T-cell polarization. Consistent with our findings in polarized T-cells, regions gaining 5hmC after one day of polarization exhibited extensive de-methylation in memory cells (P = 6.20e-78, Fisher's exact test) (**Figure 4G-I**). As 5hmC and 5mC cannot be distinguished by conventional bisulfite conversion (Nestor et al., 2014b), we performed 5hmC/5mC-specific qPCR for selected loci undergoing 5mC changes in NT, T_{CM} and T_{EM} cells (**Figure S4E-F, Table S6**).

Thus, early methylome-reprogramming events initiated by 5hmC are maintained long-term in memory T-cells *in vivo*.

5hmC remodeling marks regulatory regions enriched for disease-associated genetic variants

Consistent with previous reports in other cell types (Wu and Zhang, 2014), 5hmC in $CD4^+$ T-cells marked several different types of gene regulatory elements (such as enhancers, promoters and gene bodies). Using a Hidden Markov model together with published genome-wide data sets for several epigenetic marks in primary human NT cells (Bernstein et al., 2010; Song and Chen, 2015) (**Table S7**) we observed that 5hmC occupied a unique position in the genome, associating with both repressive and activating states, supporting the property of 5hmC as a general marker of gene regulatory activity (**Figure 5A**). Since the majority of disease-associated variants are found in non-coding regions, we compared the frequency of all disease-associated variants reported at the time of analysis ($N_{lead\text{Variant}}$ = 73,196; $P < 1x10^{-5}$, GWASdb2) (Li et al., 2012) and those in high LD (N_{VARIANT} = 600,320, $r^2 > 0.8$, 1000 Genomes Project, phase 3) in regions gaining or losing 5hmC at day 1 of *in vitro* polarization, resulting in a total of 1,560 variants (**Table S8**). Regions gaining 5hmC (hereafter referred to as "5hmC regions") were significantly enriched ($P < 0.01$) for disease associated variants in 19 diseases, where six out of the top ten diseases were autoimmune whereas regions losing 5hmC showed no significant enrichment for $CD4^+$ T-cell diseases (**Figure 5B**). Strikingly, regions gaining 5hmC were twice as likely (OR 2.6, $p < 0.0001$; Fisher's exact test) to overlap a disease-associated variant than cell type-specific enhancers (**Figure 5C**) (Hawkins et al., 2013).

5hmC has previously been associated with transcription factor binding sites (TFBS) (Yu et al., 2012) suggesting that variants in 5hmC regions may modulate gene expression by disrupting TF binding. We tested all 1,560 5hmC variants using phylogenetic module complexity analysis (PMCA), leveraging the conservation of co-occurring transcription factor binding site (TFBS) patterns within gene regulatory modules, as previously described by us

(Claussnitzer et al., 2014). Interestingly, for the variants in 5hmC regions we found 49.4% predicted to be regulatory compared to our previous findings of 33.3% in a random set of variants (Claussnitzer et al., 2014) (**Table S8**). To experimentally evaluate the effects of identified non-coding variants on regulatory protein binding we performed electrophoretic mobility-shift assays (EMSAs) using nuclear protein extracts from both transformed and primary, trait-related human $CD4^+$ T-cells, using probes for the risk and non-risk alleles of 22 variants (44 alleles) (**Table S9**). First, we tested variants associated with Crohn's disease and multiple sclerosis, as these diseases showed the highest enrichment for variants in 5hmC regions (**Figure 5B & Table S8**). We also tested variants associated with non-enriched T-cell associated diseases (such as allergy) and other non T-cell associated traits (obesity, metabolite levels) with high TFBS module conservation (**Figure 5B & Table S8**). We found an allele-specific shift in 16 out of 22 tested variants when using protein extracts from the Jurkat T-cell line (**Figure 5D** & **Figure S5A**), supporting that 5hmC can aid in the identification of regulatory variants, potentially contributing to disease pathophysiology. Importantly, variant-induced alterations on regulator binding could be replicated by using nuclear protein from primary human $CD4^+$ T-cells isolated from healthy individuals (**Figure 5E**).

As none of the risk variants affecting regulator binding were located within gene promoters or gene coding regions, we assessed how risk variants might modulate gene expression via physical interaction with distal gene regulatory elements. Using chromatin interaction data (capture Hi-C; cHi-C) for a human $CD4^+$ T-cell line (Martin et al., 2015) we found that disease-associated variants located in 5hmC regions were significantly enriched ($P_{BOOTSTRAP} < 1x10^{-4}$) for long-range chromatin interactions compared to all disease-associated variants (**Figure 5F**). Interestingly, enrichment of chromatin interactions for disease-associated variants in 5hmC regions was greater than that observed for T-cell specific enhancers (Hawkins et al., 2013) even though, not surprisingly, enhancer regions alone had a significantly higher

number of chromatin interactions compared to 5hmC regions (**Figure 5F and data not shown**). These findings strongly support the power of 5hmC as a precise and efficient marker to identify regulatory disease-associated variants.

To further validate our approach we examined interactions at the *CLEC16A* locus, which is associated with several autoimmune diseases (Wellcome Trust Case Control, 2007) (**Figure 5G**). Out of 30 variants in high LD (r^2>0.8), five were located in 5hmC regions, two of which were significantly enriched for TFBS module conservation (**Table S8**). For these two variants, we observed differential protein binding for rs7203150 in both Jurkat and primary human $CD4^+$ cells (**Figure 5D and E,** $P = 5.9 \times 10^{-7} / 1.8 \times 10^{-3}$, respectively, Student's t test), whereas no allele specific binding was observed for rs7198004 (**Table S8, Figure S5A**). The variant rs7203150 is located in intron 19 of the *CLEC16A* gene, containing many variants associated with multiple autoimmune diseases and has been shown to physically interact with and regulate the expression of a neighboring gene, *DEXI* (Davison et al., 2012). Capture Hi-C data confirmed the highly significant interaction of the rs7203150 region with the *DEXI* promoter (FDR<5%), as expected, but also revealed a novel interaction with a downstream gene, *RMI2* (**Figure 5G**). In fact, cHi-C data suggest interactions for numerous genes at the *CLEC16A* locus, *SOCS1*,a negative regulator of cytokine signaling (Diehl et al., 2000), *DEXI*, a gene of unknown function but proposed as autoimmune candidate gene (Leikfoss et al., 2013), *RMI2*, a topoisomerase critical for T-cell differentiation and a primary immunodeficiency syndrome candidate gene (Monnich et al., 2010), to *CLEC16A*, recently shown to affect T-cell selection, mediate T1D and modify CD4 single positive thymocyte reactivity (Schuster et al., 2015) (**Figure 5G**). Finally, in our T-cell data we find supportive co-expression of *CLEC16A* with ~~*CIITA*~~, *SOCS1*, *DEXI* and *RMI2* (**Figure 5G, right panel**).

Thus, 5hmC remodeling marks cell-type specific, gene regulatory regions, which may contribute to pathogenic mechanisms in $CD4^+$ T-cell associated diseases.

DISCUSSION

Previous studies of 5hmC during T-cell differentiation have reported global loss of 5hmC in differentiated T helper subsets in mouse (Ichiyama et al., 2015; Tsagaratou et al., 2014; Yang et al., 2015). We confirm these findings, but using a time-series profiling approach, we also report widespread 5hmC remodeling during early differentiation, including >10,000 locus-specific gains of 5hmC enriched at genes associated with T-cell function. Here, we report that 5hmC-mediated DNA de-methylation is a key feature of human $CD4^+$ T-cell differentiation and that tight regulation of *TET* gene expression is critical for appropriate differentiation of T-helper subsets. These DNA de-methylation events were also observed in memory T-cells *in vivo* and were enriched for variants associated with $CD4^+$ T-cell diseases including multiple sclerosis, psoriasis and Crohn's disease. Several disease-associated variants affected DNA-protein interactions in primary human $CD4^+$ T-cells and reveal novel, potentially pathogenic mechanisms in several autoimmune diseases. To our knowledge this is the first time de-methylation events have been used to trace disease-associated variants in a trait-related cell type. As 5hmC profiling can be performed on small amounts (> 25 ng) of archived, fragmented genomic DNA it is ideal for identifying and stratifying potential regulatory variants in rare primary cell types (Taiwo et al., 2012). This is in contrast to chromatin-based enhancer-profiling strategies which typically requires profiling of multiple epigenetic marks in large amounts of fresh material (Hawkins et al., 2013).

Studies in mouse have provided valuable insights into the role of DNA methylation during differentiation of mammalian $CD4^+$ T-cells, but less is known about the importance of DNA de-methylation in human $CD4^+$ T-cell differentiation and its contribution to the pathogenesis of complex immune diseases (Ichiyama et al., 2015; Ko et al., 2011; Tsagaratou et al., 2014; Yang et al., 2015). Our profiling of the methylome (5mC), hydroxymethylome

(5hmC) and transcriptome of matched human $CD4^+$ T-cells during *in vitro* differentiation of naïve T-cells (NTs) into T-helper type 1 (Th1) and T-helper type 2 (Th2) cells revealed widespread remodeling of 5hmC during early T-cell activation, and that early gains in 5hmC predicted subsequent loss of 5mC in differentiated Th1 and Th2 cells. Significantly, 5hmC remodeling was enriched in the regulatory regions of genes with known T-cell function including master regulators of lineage specification, such as *IFNG*, *TBX21* and *FOXP3*. Consistent with an important role for 5hmC mediated DNA de-methylation in T-cell differentiation, we found that (**i**) the *TET* genes are most highly expressed in T-cell related tissues ($CD4^+$ T-cells, $CD8^+$ T-cells, thymus) in humans (**ii**), $CD4^+$ T-cell activation results in dramatic (10-fold) and rapid (6 hours) down-regulation of *TET1* expression, and (**iii**) dysregulation of *TET1* expression disrupts $CD4^+$ T-cell differentiation. These observations are consistent with a growing number of studies in *Tet* knock-out mice which have revealed a pivotal role of TET enzymatic activity in T-cell biology (Ichiyama et al., 2015; Ko et al., 2010; Tsagaratou et al., 2014; Yang et al., 2015).

In addition to direct differentiation into effector subtypes, a proportion of activated NT cells become long-lived and retain 'memory' of the initial activating signal (Sallusto et al., 1999). Methylome profiling of NT, T_{CM} and T_{EM} cells *ex vivo*, revealed that the DNA methylation changes observed early during differentiation persist in memory T-cells. First, this important observation supports the validity of *in vitro* $CD4^+$ T-cell differentiation as a powerful model system in which to study the epigenetics of human $CD4^+$ T-cells. Second, these data suggest that any dysregulation of the DNA de-methylation pathway will not only affect immediate response to antigen (differentiation), but also appropriate formation of T-cell memory.

The availability of global profiles for several activating and repressive epigenetic marks allowed us to study the epigenetic neighborhood occupied by 5hmC in human $CD4^+$ T-cells.

We used a recently described Hidden Markov model based approach to identify the distinct chromatin states present in human naïve $CD4^+$ T-cells (Song and Chen, 2015). Interestingly, instead of simply following the patterning of 5-methylcytosine, from which it is derived, 5hmC was present in chromatin states composed of activating, repressive and poised histone modifications (**Figure 5A**). 5hmC's presence across so many chromatin states may reflect its property as an intermediate during conversion of transcriptionally repressive 5-methylcytosine to transcriptionally permissive unmodified cytosine (Hon et al., 2014). Alternatively, 5hmC's association with such a diverse range of histone marks may reflect the growing realization of the TET enzymes as multi-faceted proteins, connecting different layers of epigenetic and transcription control to maintain cell state (Laird et al., 2013). Indeed, the TET enzymes have been reported to directly interact with transcriptional activators including PU.1, EBF1 and p300, while also interacting with proteins associated with transcriptional repression, including EZH2, SIN3A, HDACs, NuRD and MeCP2 (Cartron et al., 2013; Delatte et al., 2014). Regardless of the underlying mechanism, 5hmC's position as a general but sensitive indicator of chromatin state make it a powerful tool in the identification of cell-type specific regulatory elements. Indeed, regions gaining 5hmC during early activation (day 1) were highly enriched for variants associated with several autoimmune diseases including Crohn's disease, multiple sclerosis, celiac disease, myasthenia gravis and psoriasis. The effect of these potentially regulatory variants on DNA-protein binding was directly tested in primary human $CD4^+$ T-cells and provides multiple molecular pathways for further investigation. By combining these results with chromatin interaction (capture Hi-C) data from a human T-cell line we found that 5hmC variants are enriched for connections between distal regulatory elements and identify several interactions, which if disrupted by risk variants, may have pathogenic consequences in autoimmune disease. Notably, dynamic 5hmC regions were more enriched for regulatory interactions than cell-type specific enhancers identified by profiling of multiple histone marks

(Hawkins et al., 2013). Combined with the ability to perform 5hmC-profiling on small amounts of archived DNA, this surprising and provocative finding suggests that 5hmC alone may be a powerful and cost-effective approach for prioritization of regulatory variants in humans.

In conclusion, 5hmC-mediated DNA de-methylation plays key roles in the differentiation of human $CD4^+$ T-cells, marking regions relevant for the pathogenesis of several autoimmune diseases and represents a novel, inexpensive and accessible approach for identifying causal disease variants in autoimmune disease.

EXPERIMENTAL PROCEDURES

Ethics Statement

This study was approved by the ethics board of Linkoping University and all participants provided written consent for participation.

Cell Isolation and stimulation

Peripheral blood mononuclear cells (PBMCs) were enriched from healthy donor buffy coats using Lymphoprep (Axis-shield). Human total $CD4^+$ T-cells or Naïve $CD4^+$ T-cells were then isolated through magnetic sorting (Miltenyi Biotec). Naïve T-cells were cultured with plate bound anti-CD3 (500ng/µL) and soluble anti-CD28 (500ng/µL) in the presence of IL-12 (5ng/mL), IL-2 (10ng/mL) and anti-IL4 (5µg/mL) for Th1- or IL-4 (10ng/mL), IL-2 (10ng/mL), anti-IFNG (5µg/mL) and anti-IL12 (5µg/mL) for Th2 conditions. Cells were grown for up to five days in RPMI-1640 medium (Life technologies) supplemented with 2mM L-glutamine (PAA laboratories), 10% heat-inactivated FCS (PAA laboratories) and 50µg/mL gentamicin (Sigma-Aldrich). Cells were re-stimulated with respective condition (see above) after three days in culture.

DNA and RNA extraction

Total RNA and genomic DNA was extracted using AllPrep DNA/RNA mini kit (Qiagen). DNA and RNA integrity was determined using a 2100 Bioanalyzer (Agilent).

hMeDIP-Seq and data analysis

DNA (1.5µg) was fragmented by sonication using a Bioruptor (6 x 15min 30s on/off, Diagenode) then subjected to end-repair, dA-tailing, and adaptor-ligation using the NEBNext DNA Library Prep Master Mix Set for Illumina (New England Biolabs). Samples were hydroxymethylated DNA immuno-precipitated (hMeDIP) as previously described (Nestor et al., 2012). Briefly, DNA was denatured and incubated with 1µg antibody against 5hmC (ActiveMotif; #39769) overnight at 4°C. IP and input DNA was prepared using magnetic beads (Dynabeads Protein G; Invitrogen) and amplified using NEBNext Multiplex Oligos for Illumina (New England Biolabs). Samples were separated by electrophoresis on a 2% low melting point agarose gel and fragments between 100 – 400 bp were selected. DNA was purified using the QIAquick Gel Extraction Kit (QIAGEN) and sequenced on an Illumina HiSeq 2000 platform. Reads were mapped to the human genome (hg19) using bowtie with the following parameters (bowtie –v 1 – best -S). Peak calling was performed using MACS with input samples as control.

Immuno dot-blotting

DNA was denatured and applied to a positively charged nylon membrane under vacuum using a Dot Blot Hybridisation Manifold (Harvard Apparatus Limited, UK). The membrane was washed twice in 2X SSC buffer, air-dried and UV-crosslinked. Membranes were incubated with an antibody against 5hmC (1:3000, Active Motif) for 1h at 4°C then washed in TBS-

Tween (0.05%) and incubated with a HRP conjugated goat-anti-rabbit antibody (1:10000, Bio-Rad). Following treatment with enhanced chemiluminescence (ECL) substrate, membranes were scanned on a ChemiDoc MP imaging system (Bio-Rad). To control for loading, membranes were stained with methylene blue. Spot intensities were quantified using ImageJ (NIH).

Gene expression microarrays and analysis

For gene expression microarrays, RNA was labeled and amplified using the Low Input Quick Amp Labeling kit (Agilent Technologies), then hybridized onto SurePrint G3 Human Gene Expression 8x60K v2 microarrays (Agilent Technologies) and scanned using a Surescan High Resolution DNA Microarray Scanner (Agilent Technologies). Raw intensities were exported with Agilent's Feature Extraction Software. All subsequent analyses were performed using the LIMMA package in the R statistical programming language. Briefly, data was background corrected and quantile normalized then control probes, probes not expressed (background +10%) in all conditions and non-annotated probes were removed.

Statistical analysis

P-values < 0.05 were considered significant and when stated, p-values were adjusted using Benjamini-Hochberg correction.

Data access

All raw and processed microarray and next-generation sequencing data will be made publically available in the ArrayExpress database.

AUTHOR CONTRIBUTIONS

C.E.N. and M.B. conceived project, designed experiments and wrote the paper. C.E.N., A.L., H.Z., C.H.N., H.W., O.R., L.M., B.K. and D.G. performed experiments and data analysis. H.L., S.M.H, M.G. and M.S designed experiments and wrote the paper. R.R.M. conceived experiments and wrote the paper.

ACKNOWLEDGEMENTS

M.S. and B.K. are employees of Genomatix Software GmBH. The authors declare no conflicts of interest.

SUPPLEMENTAL INFORMATION

Supplemental information contains five figures, six tables, and Supplemental Experimental Procedures.

REFERENCES

Bernstein, B.E., Stamatoyannopoulos, J.A., Costello, J.F., Ren, B., Milosavljevic, A., Meissner, A., Kellis, M., Marra, M.A., Beaudet, A.L., Ecker, J.R., *et al.* (2010). The NIH Roadmap Epigenomics Mapping Consortium. Nature biotechnology *28*, 1045-1048.

Cartron, P.F., Nadaradjane, A., Lepape, F., Lalier, L., Gardie, B., and Vallette, F.M. (2013). Identification of TET1 Partners That Control Its DNA-Demethylating Function. Genes & cancer *4*, 235-241.

Claussnitzer, M., Dankel, S.N., Klocke, B., Grallert, H., Glunk, V., Berulava, T., Lee, H., Oskolkov, N., Fadista, J., Ehlers, K., *et al.* (2014). Leveraging cross-species transcription factor binding site patterns: from diabetes risk loci to disease mechanisms. Cell *156*, 343-358.

Davison, L.J., Wallace, C., Cooper, J.D., Cope, N.F., Wilson, N.K., Smyth, D.J., Howson, J.M., Saleh, N., Al-Jeffery, A., Angus, K.L., *et al.* (2012). Long-range DNA looping and gene expression analyses identify DEXI as an autoimmune disease candidate gene. Human molecular genetics *21*, 322-333.

Delatte, B., Deplus, R., and Fuks, F. (2014). Playing TETris with DNA modifications. The EMBO journal *33*, 1198-1211.

Diehl, S., Anguita, J., Hoffmeyer, A., Zapton, T., Ihle, J.N., Fikrig, E., and Rincon, M. (2000). Inhibition of Th1 differentiation by IL-6 is mediated by SOCS1. Immunity *13*, 805-815.

Graves, M., Benton, M., Lea, R., Boyle, M., Tajouri, L., Macartney-Coxson, D., Scott, R., and Lechner-Scott, J. (2013). Methylation differences at the HLA-DRB1 locus in CD4+ T-Cells are associated with multiple sclerosis. Multiple sclerosis *20*, 1033-1041.

Guo, J.U., Su, Y., Zhong, C., Ming, G.L., and Song, H. (2011). Hydroxylation of 5-methylcytosine by TET1 promotes active DNA demethylation in the adult brain. Cell *145*, 423-434.

Gustafsson, M., Gawel, D.R., Alfredsson, L., Baranzini, S., Bjorkander, J., Blomgran, R., Hellberg, S., Eklund, D., Ernerudh, J., Kockum, I., *et al.* (2015). A validated gene regulatory network and GWAS identifies early regulators of T cell-associated diseases. Science translational medicine *7*, 313ra178.

Hawkins, R.D., Larjo, A., Tripathi, S.K., Wagner, U., Luu, Y., Lonnberg, T., Raghav, S.K., Lee, L.K., Lund, R., Ren, B., *et al.* (2013). Global chromatin state analysis reveals lineage-specific enhancers during the initiation of human T helper 1 and T helper 2 cell polarization. Immunity *38*, 1271-1284.

Hon, G.C., Song, C.X., Du, T., Jin, F., Selvaraj, S., Lee, A.Y., Yen, C.A., Ye, Z., Mao, S.Q., Wang, B.A., *et al.* (2014). 5mC Oxidation by Tet2 Modulates Enhancer Activity and Timing of Transcriptome Reprogramming during Differentiation. Molecular cell *56*, 286-297.

Ichiyama, K., Chen, T., Wang, X., Yan, X., Kim, B.S., Tanaka, S., Ndiaye-Lobry, D., Deng, Y., Zou, Y., Zheng, P., *et al.* (2015). The methylcytosine dioxygenase tet2 promotes DNA demethylation and activation of cytokine gene expression in T cells. Immunity *42*, 613-626.

Janson, P.C., Linton, L.B., Bergman, E.A., Marits, P., Eberhardson, M., Piehl, F., Malmstrom, V., and Winqvist, O. (2011). Profiling of CD4+ T cells with epigenetic immune lineage analysis. Journal of immunology *186*, 92-102.

Ko, M., Bandukwala, H.S., An, J., Lamperti, E.D., Thompson, E.C., Hastie, R., Tsangaratou, A., Rajewsky, K., Koralov, S.B., and Rao, A. (2011). Ten-Eleven-Translocation 2 (TET2) negatively regulates homeostasis and differentiation of hematopoietic stem cells in mice. Proceedings of the National Academy of Sciences of the United States of America *108*, 14566-14571.

Ko, M., Huang, Y., Jankowska, A.M., Pape, U.J., Tahiliani, M., Bandukwala, H.S., An, J., Lamperti, E.D., Koh, K.P., Ganetzky, R., *et al.* (2010). Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant TET2. Nature *468*, 839-843.

Laird, A., Thomson, J.P., Harrison, D.J., and Meehan, R.R. (2013). 5-hydroxymethylcytosine profiling as an indicator of cellular state. Epigenomics *5*, 655-669.

Lee, G.R., Kim, S.T., Spilianakis, C.G., Fields, P.E., and Flavell, R.A. (2006). T helper cell differentiation: regulation by cis elements and epigenetics. Immunity *24*, 369-379.

Leikfoss, I.S., Mero, I.L., Dahle, M.K., Lie, B.A., Harbo, H.F., Spurkland, A., and Berge, T. (2013). Multiple sclerosis-associated single-nucleotide polymorphisms in CLEC16A correlate with reduced SOCS1 and DEXI expression in the thymus. Genes and immunity *14*, 62-66.

Li, M.J., Wang, P., Liu, X., Lim, E.L., Wang, Z., Yeager, M., Wong, M.P., Sham, P.C., Chanock, S.J., and Wang, J. (2012). GWASdb: a database for human genetic variants identified by genome-wide association studies. Nucleic acids research *40*, D1047-1054.

Li, Z., Cai, X., Cai, C.L., Wang, J., Zhang, W., Petersen, B.E., Yang, F.C., and Xu, M. (2011). Deletion of Tet2 in mice leads to dysregulated hematopoietic stem cells and subsequent development of myeloid malignancies. Blood *118*, 4509-4518.

Licona-Limon, P., Kim, L.K., Palm, N.W., and Flavell, R.A. (2013). TH2, allergy and group 2 innate lymphoid cells. Nature immunology *14*, 536-542.

Liu, Y., Aryee, M.J., Padyukov, L., Fallin, M.D., Hesselberg, E., Runarsson, A., Reinius, L., Acevedo, N., Taub, M., Ronninger, M., *et al.* (2013). Epigenome-wide association data implicate DNA methylation as an intermediary of genetic risk in rheumatoid arthritis. Nature biotechnology *31*, 142-147.

Lu, F., Liu, Y., Jiang, L., Yamaguchi, S., and Zhang, Y. (2014). Role of Tet proteins in enhancer activity and telomere elongation. Genes & development *28*, 2103-2119.

Martin, P., McGovern, A., Orozco, G., Duffus, K., Yarwood, A., Schoenfelder, S., Cooper, N.J., Barton, A., Wallace, C., Fraser, P., *et al.* (2015). Capture Hi-C reveals novel candidate genes and complex long-range interactions with related autoimmune risk loci. Nature communications *6*, 10069.

Monnich, M., Hess, I., Wiest, W., Bachrati, C., Hickson, I.D., Schorpp, M., and Boehm, T. (2010). Developing T lymphocytes are uniquely sensitive to a lack of topoisomerase III alpha. European journal of immunology *40*, 2379-2384.
Nestor, C.E., Barrenas, F., Wang, H., Lentini, A., Zhang, H., Bruhn, S., Jornsten, R., Langston, M.A., Rogers, G., Gustafsson, M., and Benson, M. (2014a). DNA methylation changes separate allergic patients from healthy controls and may reflect altered CD4+ T-cell population structure. PLoS genetics *10*, e1004059.
Nestor, C.E., Ottaviano, R., Reddington, J., Sproul, D., Reinhardt, D., Dunican, D., Katz, E., Dixon, J.M., Harrison, D.J., and Meehan, R.R. (2012). Tissue type is a major modifier of the 5-hydroxymethylcytosine content of human genes. Genome research *22*, 467-477.
Nestor, C.E., Ottaviano, R., Reinhardt, D., Cruickshanks, H.A., Mjoseng, H.K., McPherson, R.C., Lentini, A., Thomson, J.P., Dunican, D.S., Pennings, S., *et al.* (2015). Rapid reprogramming of epigenetic and transcriptional profiles in mammalian culture systems. Genome biology *16*, 11.
Nestor, C.E., Reddington, J.P., Benson, M., and Meehan, R.R. (2014b). Investigating 5-hydroxymethylcytosine (5hmC): the state of the art. Methods in molecular biology *1094*, 243-258.
Sallusto, F., Geginat, J., and Lanzavecchia, A. (2004). Central memory and effector memory T cell subsets: function, generation, and maintenance. Annual review of immunology *22*, 745-763.
Sallusto, F., Lenig, D., Forster, R., Lipp, M., and Lanzavecchia, A. (1999). Two subsets of memory T lymphocytes with distinct homing potentials and effector functions. Nature *401*, 708-712.
Schuster, C., Gerold, K.D., Schober, K., Probst, L., Boerner, K., Kim, M.J., Ruckdeschel, A., Serwold, T., and Kissler, S. (2015). The Autoimmunity-Associated Gene CLEC16A Modulates Thymic Epithelial Cell Autophagy and Alters T Cell Selection. Immunity *42*, 942-952.
Song, C.X., Szulwach, K.E., Fu, Y., Dai, Q., Yi, C., Li, X., Li, Y., Chen, C.H., Zhang, W., Jian, X., *et al.* (2011). Selective chemical labeling reveals the genome-wide distribution of 5-hydroxymethylcytosine. Nature biotechnology *29*, 68-72.
Song, J., and Chen, K.C. (2015). Spectacle: fast chromatin state annotation using spectral learning. Genome biology *16*, 33.
Tahiliani, M., Koh, K.P., Shen, Y., Pastor, W.A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L.M., Liu, D.R., Aravind, L., and Rao, A. (2009). Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. Science *324*, 930-935.
Taiwo, O., Wilson, G.A., Morris, T., Seisenberger, S., Reik, W., Pearce, D., Beck, S., and Butcher, L.M. (2012). Methylome analysis using MeDIP-seq with low DNA concentrations. Nature protocols *7*, 617-636.
Tsagaratou, A., Aijo, T., Lio, C.W., Yue, X., Huang, Y., Jacobsen, S.E., Lahdesmaki, H., and Rao, A. (2014). Dissecting the dynamic changes of 5-hydroxymethylcytosine in T-cell development and differentiation. Proceedings of the National Academy of Sciences of the United States of America *111*, E3306-3315.
Wahren-Herlenius, M., and Dorner, T. (2013). Immunopathogenic mechanisms of systemic autoimmune disease. Lancet *382*, 819-831.
Wellcome Trust Case Control, C. (2007). Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature *447*, 661-678.
Williams, K., Christensen, J., Pedersen, M.T., Johansen, J.V., Cloos, P.A., Rappsilber, J., and Helin, K. (2011). TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. Nature *473*, 343-348.

Wu, H., D'Alessio, A.C., Ito, S., Xia, K., Wang, Z., Cui, K., Zhao, K., Sun, Y.E., and Zhang, Y. (2011). Dual functions of Tet1 in transcriptional regulation in mouse embryonic stem cells. Nature *473*, 389-393.
Wu, H., and Zhang, Y. (2014). Reversing DNA methylation: mechanisms, genomics, and biological functions. Cell *156*, 45-68.
Yamane, H., and Paul, W.E. (2013). Early signaling events that underlie fate decisions of naive CD4(+) T cells toward distinct T-helper cell subsets. Immunol Rev *252*, 12-23.
Yang, R., Qu, C., Zhou, Y., Konkel, J.E., Shi, S., Liu, Y., Chen, C., Liu, S., Liu, D., Chen, Y., *et al.* (2015). Hydrogen Sulfide Promotes Tet1- and Tet2-Mediated Foxp3 Demethylation to Drive Regulatory T Cell Differentiation and Maintain Immune Homeostasis. Immunity *43*, 251-263.
Yu, M., Hon, G.C., Szulwach, K.E., Song, C.X., Zhang, L., Kim, A., Li, X., Dai, Q., Shen, Y., Park, B., *et al.* (2012). Base-resolution analysis of 5-hydroxymethylcytosine in the mammalian genome. Cell *149*, 1368-1380.

FIGURE LEGENDS

Figure 1. Dynamic remodeling of 5-hydroxymethylcytosine (5hmC) during *in vitro* polarization of human CD4+ T-cells.

(**A**) Schematic figure of experimental design for T-cell polarization.

(**B**) Global 5hmC content measured by immuno-dot blot using a 5hmC antibody. Naïve T-cells (NT) were cultured under Th1 or Th2 polarizing conditions for up to 5 days. Brain is shown as a tissue with high 5hmC content. Data shown as mean ±SD, representative of 5 biological replicates. *p<0.05 Student's t test.

(**C**) Normalized 5hmC density profiles across gene body ± 4 kb flanking regions.

(**D**) Normalized 5hmC density profiles across gene body ± 2 kb flanking regions in NT cells binned into three equal sized groups based on gene expression levels.

(**E**) Genomic distribution of 5hmC peaks during T-cell differentiation *in vitro*. ***p<0.001 Fisher's Exact test.

(**F**) Gene ontology (GO) enrichment of 5hmC peaks in T-cells differentiated *in vitro* for 1 day.

(**G**) Coverage plots of 5hmC levels at *IFNG*, *IL4*, *IL5*, *AIM2* and *IL23R* loci showing subset specific changes.

See also Figure S1

Figure 2. DNA de-methylation occurs via 5-hydroxymethylcytosine (5hmC) during *in vitro* differentiation of human CD4$^+$ T-cells.

(**A**) Volcano plot showing changes in DNA methylation (5mC) during *in vitro* polarization of CD4$^+$ T-cells relative to Naïve T-cells (NT). Vertical lines indicates a change of 20%. 5mC measured by 450K methylation array.

(**B**) 5hmC regions show major loss of 5mC during T-cell polarization. Background represents a randomly sampled group with the same size of probes in 5hmC peaks. Fisher's exact test was used to calculate significance (rightmost).

(**C**) Line plot (top) showing regions getting hypomethylated typically gain 5hmC at day 1 of T cell polarization. Box plot (bottom) showing genes associated with regions becoming hypermethylated decrease in expression levels. ***p<0.001 one-way ANOVA Tukey's test.

(**D**) Coverage plot showing active enhancers (H3K4me1, H3K27ac) gaining 5hmC and losing 5mC during T-cell polarization. Representative loci of *AIM2* and *CCL5* shown.

(**E**) Bar chart of 5mC changes in T-cell specific enhancer regions (H3K4me1, H3K27ac) show higher degree of remodeling in these regions compared to genome-wide. 5mC was measured by 450K methylation array.

(**F,G**) Distribution of 5mC changes at T-cell specific enhancers (H3K4me1, H3K27ac) show an early loss of 5mC in these regions.

See also Figure S2

Figure 3. Tight regulation of *TET* gene expression is required for appropriate T-cell differentiation

(**A**) Barplot of *TET1* gene expression in healthy human primary tissues (pool of at least 5 individuals) and primary immune cell subsets analyzed by qPCR. NT: Naïve T-cell.

(**B, C**) Barplots of *TET* gene expression measured by qPCR in Naïve T-cells (NT) cultured under Th1 or Th2 polarizing conditions. Expression levels shown relative to *GUSB*. Data shown as mean ±SD, representative of 3 biological replicates. *p<0.05, **p<0.01 Student's t test.

(**D**) Schematic figure of plasmids containing full length *TET1* (TET1fl), catalytic domain of *TET1* (TET1cd) and mutated catalytic domain of *TET1* (TET1mut).

(**E, F**) Barplots of gene expression for selected genes in NT cells transfected with TET1 plasmids then cultured under Th1 polarizing conditions for 24h. Lines indicate a log2 fold change of 0.5. Gene expression measured by microarray.

See also Figure S3

Figure 4. Early DNA methylation remodeling persists in CD4$^+$ memory T-cells *in vivo*.

(**A**) Barplot of *TET* gene expression in primary CD4$^+$Naïve T-cells (NT), central memory (TCM) and effector memory (TEM) cells. Gene expression of *TET1/2/3* was measured by qPCR.

(**B**) Barplot of global 5hmC content measured by immuno-dot blot using a 5hmC antibody in primary NT, TCM and TEM cells.

(**C**) Heatmap of gene expression in primary NT, TCM and TEM cells showing subset specific gene signatures. Gene expression measured by microarray.

(**D**) Unsupervised hierarchical clustering of DNA methylation (5mC) in NT, TCM and TEM cells measured by 450K methylation array. Significance calculated by bootstrap resampling.

(**E**) Volcano plot of 5mC changes in memory subsets showing a predominant loss in both TCM and TEM cells. Vertical lines indicate a change of 30%.

(**F**) Correlation between 5mC and gene expression in TEM cells calculated using Spearman's rank correlation coefficient. 5mC measured by 450k array and gene expression measured by microarray.

(**G**) 5hmC regions show major loss of 5mC in TEM cells. Background represents a randomly sampled group with the same size of probes in 5hmC peaks. Significance calculated using Fisher's exact test (rightmost).

(**H**) Venn diagram of sites losing 5mC in TEM cells and after 5 days of polarization towards Th1 and Th2. Sites losing 5mC defined as a loss of 20% or 30% 5mC vs. NT for polarization and TEM, respectively, and $p<0.05$. P-value for overlaps calculated using Fisher exact test.

(**I**) Coverage plot of 5mC in NT,TCM and TEM cells showing loss at representative loci marked by 5hmC during *in vitro* polarization of T-cells. ***$p<0.001$ Student's t test.

(**A**,**B**) Data shown as mean ±SD of at least 3 biological replicates. *$p<0.05$, **$p<0.01$, ***$p<0.001$, Student's t test.

Figure 5. 5hmC regions are highly enriched for disease-associated variants.

(**A**) ChromHMM heatmap of enrichment and colocalization of 5hmC with other epigenetic marks in Naïve $CD4^+$T-cells.

(**B**) Enrichment of disease-associated variants in peaks gaining- (left) and losing 5hmc (right) after 1 day of $CD4^+$T-cell polarization.

(**C**) Venn diagram of disease-associated variant localization in 5hmC 1 day gain peaks and CD4+T-cell enhancers.

(**D,E**) EMSA analysis of protein-DNA binding showing changes in binding upon introduction of disease-associated variant~~SNP~~s in 5hmC regions using nuclear protein extracts from Jurkat T-cell line (**D**) and primary $CD4^+$T-cells (**E**). Dashed boxes indicate shifts in binding, arrows indicate shifts not observed in cell line.

(**F**) Barplot of capture Hi-C (cHi-C) interactions overlapping variants in identified 5hmC regions or T-cell enhancers. Region sizes were normalized to avoid size bias and P-values calculated using bootstrap resampling n=10,000.

(**G**) Genomic plot of variant rs7203150 located in identified 5hmC region showing interactions with nearby gene promoters (left). Heatmap of co-expression during T-cell polarization showing high degree of co-expression between genes interacting with variant rs7203150 (right). cHi-C: capture Hi-C.

See also Figure S5

Figure 1 (Nestor & Lentini et al.)

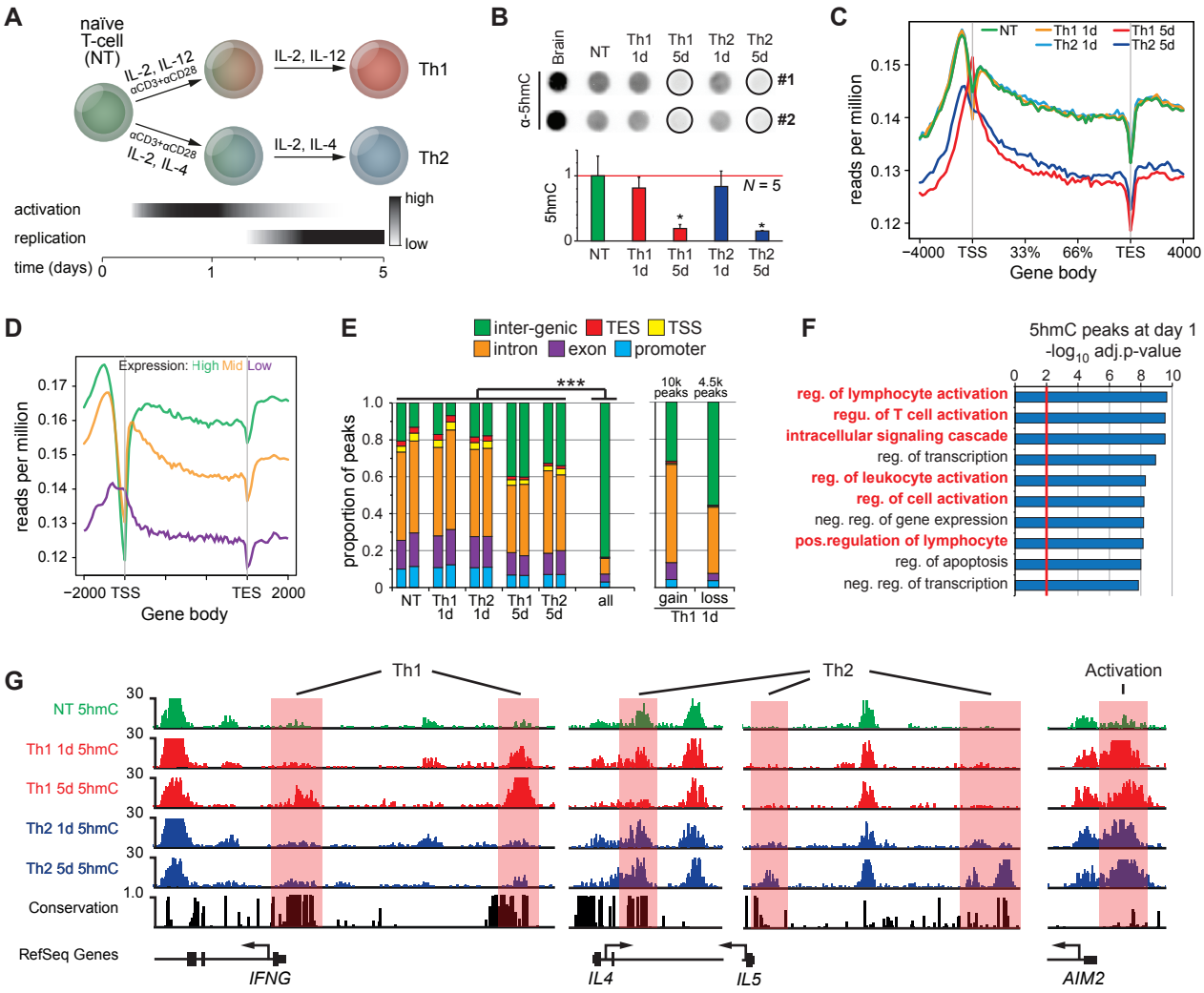

Figure 2 (Nestor & Lentini et al.)

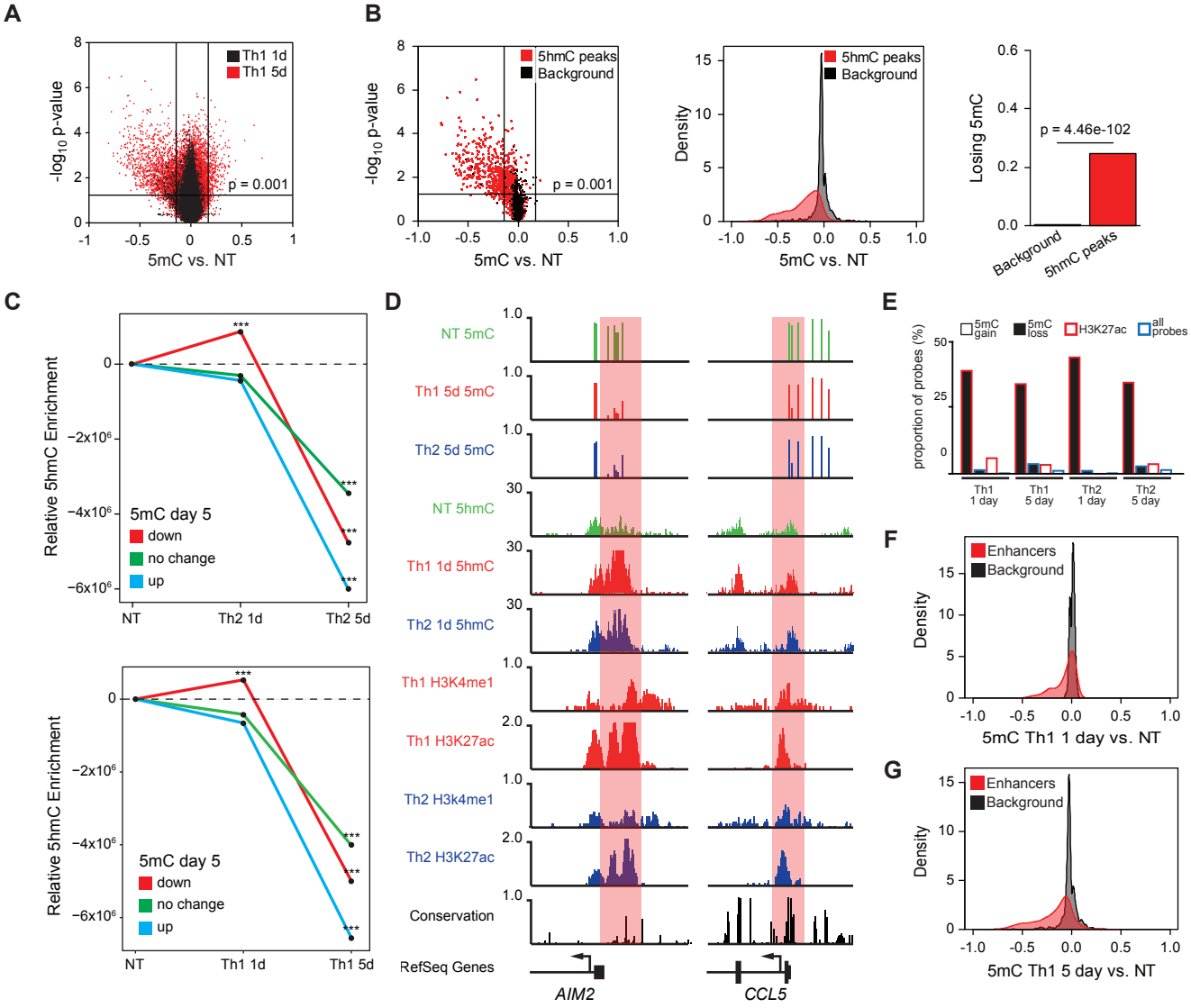

Figure 3 (Nestor & Lentini et al.)

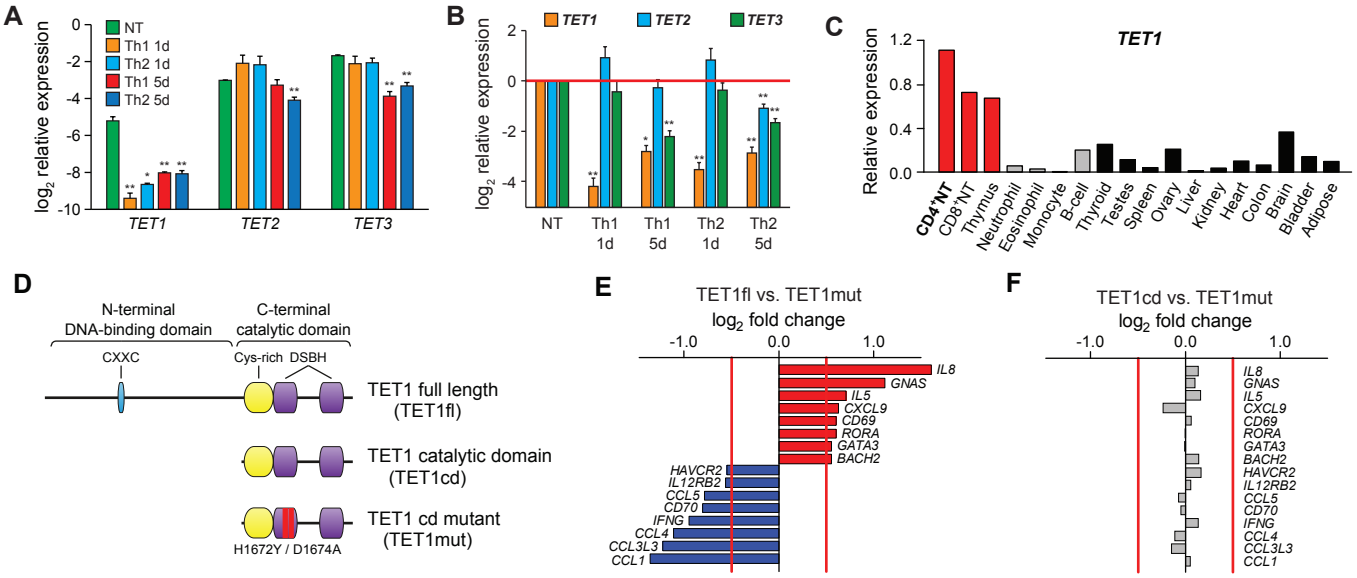

Figure 4 (Nestor & Lentini et al.)

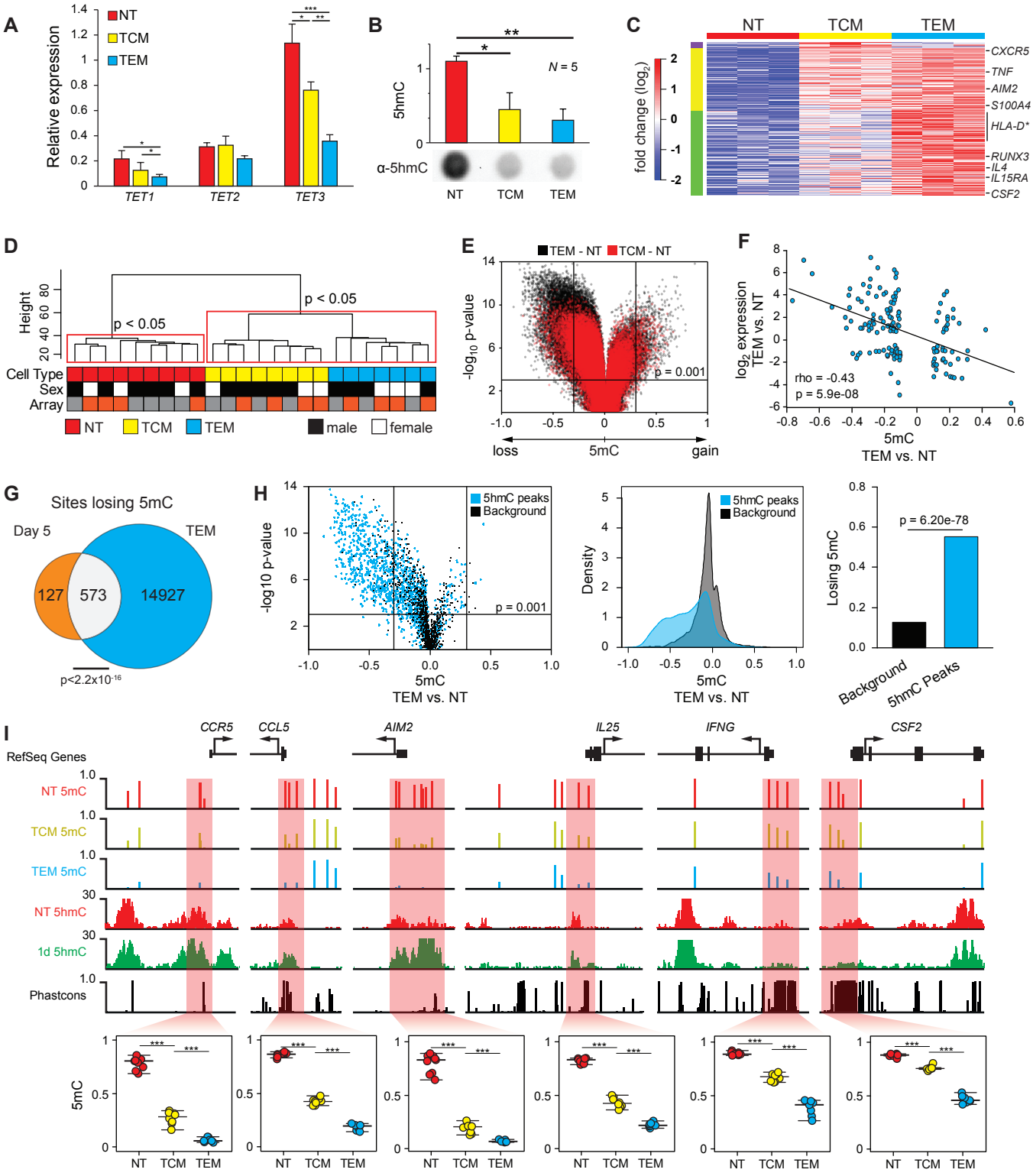

Figure 5 (Nestor & Lentini et al.)